DOCUMENT RESUME

ED 294 931 TM 011 701

AUTHOR Fein, Susan; Solomon, Alan
TITLE Nine-Year-Old Hispanic Student Writing Performance: A Replication of the 1984-85 National Assessment of Educational Progress.
PUB DATE Apr 88
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary School Students; *Grade 4; *Hispanic Americans; Intermediate Grades; Persuasive Discourse; Sex Differences; Writing (Composition); *Writing Evaluation
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

An attempt was made to replicate findings of the National Assessment of Educational Progress (NAEP) among nine-year-old Hispanic students in Grade 4. The subjects were from 12 classrooms in four elementary schools with high Hispanic student enrollments. All writing activities took place within the respective classrooms during the morning, and the students were given as much time to complete the writing assignments as they needed. Ninety-two imaginative writing samples were collected; 72 informative writing samples were collected; and 76 persuasive writing samples were collected. Results were compared to NAEP findings for the national sample. For informative and persuasive writing, the NAEP sample produced superior writing samples, but the study sample was superior in imaginative writing. Within the study sample, no significant differences were attributable to gender for imaginative or informative writing, but females were superior to males in persuasive writing. The distribution of scores in the study sample was different from NAEP findings, indicating that the NAEP results have applications as a criterion in schools and do not exist for their own sake. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294931

NINE YEAR-OLD HISPANIC STUDENT WRITING PERFORMANCE: A REPLICATION OF THE 1984-85 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

SUSAN FEIN

ALAN SOLOMON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Prepared by,

Susan Fein

Supervisor of Reading, The Abington Schools

and

Alan Solomon

Research Associate, The School District of Philadelphia

The American Educational Research Association

New Orleans, LA

April 5, 1988

TM 011 701

BEST COPY AVAILABLE

Nine Year-Old Hispanic Student Writing Performance: A Replication of the 1984-85 National Assessment of Educational Progress

The 1984-85 National Assessment of Educational Progress (NAEP) examined informative, persuasive and imaginative writing among black, white and Hispanic students at three ages, nine, thirteen and seventeen. Gender and geographic region were also used as categorical variables in the study. While many studies have focused on writing, little research has been conducted among Hispanics. Moreover, Applebee et al. (1986) suggested that the percent of Hispanic students in the NAEP study was too small to interpret without caution.

Hispanics tend to congregate in certain areas of the country and their representation in the school districts which serve them may exceed the sample proportions. In Philadelphia, for instance, one of the School District's administrative units has a Hispanic student enrollment which roughly triples the NAEP sample for nine year-olds. In this unit, Hispanic students accounted for 34.9 percent of the fourth grade student enrollment where a majority of the students are nine years old. In the NAEP sample, the proportion of Hispanic students ranged from 7 percent to12 percent. Therefore, while the NAEP has performed a valuable service by addressing Hispanic student writing performance and pointing out directions for researchers to follow, verification is necessary in order to support the study's findings.

In order to measure writing, the NAEP used a prompt and evaluated the samples holistically and by primary trait scoring (PTS). Because of time restrictions, the researchers used only PTS, a scoring system based on the assumption that different assignments must be judged on different criteria (Odell, 1981). Criteria have to be set in terms of the audience's characteristics by rhetorical statements which reflect the writer's ability. When evaluating a persuasive writing exercise, for instance, the evaluator sets criteria designed to determine if the writer identified the problem, prepared a solution to it and demonstrated that the solution was workable and beneficial. Judgments based on these criteria can be used to form summative evaluations of students' writing and to generate data for research activities and curriculum evaluations (Cooper and Odell, 1977).

There are some problems associated with PTS but they can be resolved with relatively little difficulty. Primarily, the scoring procedure does not ask evaluators to examine textual issues such as

cohesion. To address this concern, the NAEP supplemented PTS with error analysis, syntax checks and coherence. Second, PTS restricts the issues judges may consider when they evaluate writing samples. Thus, judges tend to discount responses which go beyond the task's parameters or work through an unanticipated perspective in order to deal with the matter at hand. Odell suggested that the second problem could be resolved by identifying unusual selections and grading them separately. Additionally, evaluators can set range finders in their criteria. Odell also claimed that PTS is a sound procedure for combining diagnosis and evaluation.

White (1986) credited Lloyd-Jones with providing the best available summary of PTS's conceptual history. Lloyd-Jones was a member of the group convened by NAEP to work on the matter and, "exemplifies the wit, sensitivity, and pedagogical experience that were part of the entire enterprise" (p. 143). Lloyd-Jones (1977) stated that writing and discourse were synonymous and samples should be examined in line with their aims and features. Aims are linked to the functions of language and features, to its mechanics. While judgments on writing quality are based on aims, precise issues are rooted in features.

In themselves, writing assessments may be atomistic or holistic with both types having some advantages. Lloyd-Jones said that atomistic tests are more reliable and holistic measures, more valid. Of the available holistic measures, PTS provides the most meaningful information. "The goal of Primary Trait Scoring is to define precisely what segment of discourse will be evaluated (e. g., presenting rational persuasion between social equals in a formal situation), and to train readers to render holistic judgments accordingly" (p. 37). To this end, PTS users have to define their universe, prepare appropriate exercises, ensure the writers' cooperation, and prepare scoring guides.

The major problem associated with holistic scoring procedures emerges when scoring guides are either too general or not pertinent. Therefore, the group which developed PTS examined the history of rhetorical theory to generate a means of focusing assessment based on a "consistent understanding of the goals of writing" (p. 143). The team produced a three part scoring strategy. This strategy included expressive, explanatory and persuasive modes which generated a set of exercises and scoring techniques designed to produce information about the writing samples studied.

White contended that the advantage of PTS in classroom situations is obvious in that it allows the teacher to focus on one issue. When writing's surface features are not important, PTS becomes a

scoring method without wide scope. Thus, PTS allows teachers to concentrate their efforts in writing instruction.

Fuller (1985) used the case study technique in her investigation of PTS. Student compositions were read and rated for their effectiveness. A descriptive scoring guide which evolved from the papers was used for the rating procedure. Fuller analyzed the papers in order to determine if the raters could ignore secondary writing traits while they were evaluating the papers. The raters scored the same papers twice for reliability and the second session was recorded to study the interactions between the raters and their ramifications. Fuller questioned the validity of PTS because her findings revealed that the scores represented the interaction of the text, the scoring guide and the social setting as well as the writing sample evaluation.

Farmer (1986) studied the relationship between large-scale assessments of educational proficiency with instructional practice and the processes students use when learning how to write. The investigator dealt with the reliability and validity of PTS. According to Farmer, researchers have shifted their attention from the finished written product to the strategies students employ when they write. Here, students learn writing by working through a complex recursive process which includes a series of prewriting activities, the preparation of rough drafts, revision cycles and a final draft. Assessments have ignored these developments and directed their attention toward the production of an impromptu written response with artificial time constraints which preclude prewriting or revisions.

In this study, the researcher tried to determine if allowing students to work through the steps of the writing process in a test influenced the scores and their reliability and validity. Farmer used two approaches in her study, traditional and process. Thirty-six fourth grade classes joined the study. Students were randomly assigned to treatment groups. A writing sample was given with a prompt used in the1984 NAEP writing examination. Farmer found that writing quality was high for both groups but the process cohort mean was significantly higher than that of the traditional cohort. Interrater reliablity and concurrent validity were low in both instances.

Swartz (1986) studied the variability of fourth grade student writing performance through three compositions. The investigator took steps to answer three questions dealing with the reliability of direct writing assessment with fourth grade children. Swartz's study is pertinent here because children at this level served as the study group for this investigation.

Swartz asked three questions in her study. First, how does rater variability compare to writer variability ? Second, what are the relationships between reliability estimates for different writing elements ? Third, how many raters and samples are necessary to confirm the reliability of measurement for fourth grade writing skills ?

In the study, 120 fourth grade students were asked to write narrative compositions once a week for three weeks. Pictures prepared by the NAEP were used as prompts. The essays were scored holistically by four judges. The judges also scored the essays in four analytic categories: (1) organization, (2) language, (3) sentence structure and (4) a combination of capitalization and punctuation.

Swartz estimated the variance components for students, raters, topics and topic sequence from analysis of variance. The first or second greatest source of variability in writing on all scoring categories, aside from language, was the individual student. Organization skills showed the most variability and sentence structure, the least.

Swartz used generalizability theory to produce reliability estimates. The investigator developed information for one to four samples and two to six raters. To produce reliability estimates in excess of .60 for holistic scoring, Swartz claimed that at least two raters and two samples are necessary. Two additional raters are necessary to achieve this level for language and sentence structure skills. For organization, capitalization and punctuation, three samples are necessary because of variation across writing samples.

Mitchell and Anderson (1986) conducted a reliability study on the holistic scoring procedure. The researchers studied a sample of essays written by a group of examinees who took the spring, 1985 Medical College Admission Test. Through their essay, the examinees could show their skills in six areas: (1) developing a central idea, (2) synthesizing concepts and ideas, (3) identifying relevant and irrelevant information, (4) forming alternative hypotheses, (5) presenting ideas cohesively and logically, and (6) writing clearly.

Twenty raters scored 3,117 papers. Groups of twenty essays were prepared and each was assigned at random to two raters. Essays were rated on a six point holistic scale. The ratings were checked for agreement and those in which the raters differed by more than one point were read by a

third rater. A third reading took place for 5.3 percent of the sample. This finding revealed that the raters were of a similar mind in terms of evaluating the essays they read.

Two hundred and seventy-nine essays were read by four raters in order to produce reliability data. Time, batch size and reading group appeared to influence scoring. The mean for the second reading day was farther from the six point holistic scale mean than the means for the first and third days. Although scores from one batch to another differed appreciably, the writers were unable to determine the reason for this phenomenon. They suggested using smaller batches in order to control this problem. Leadership seemed to influence the scores produced by each reading group. Rotating leaders among groups may resolve this problem.

Isern (1986) examined bias in Hispanic writing performance. The researcher used more than 2,800 first-year college students as her sample and worked with both direct and indirect measures, an objective writing test and a holistic sample. Isern found some bias in scoring the writing sample. "However, the prediction equation of the majority group overpredicted the performance of the minority group in the English course by such a small amount, that the statistical significance could be attributed to sample size" (p. 2135). Consequently, the researcher found limited, if any, bias in this assessment. This procedure seems to be applicable for elementary students as well.

Casillas (1986) studied the relationships between writing interests, selected writing traits, and reading and writing scores among fourth and sixth grade Hispanic students. The researcher conducted her study in Texas and used the Writing Interest Inventory, the Writing Traits Scale and the reading and writing scores from the Texas Assessment of Basic Skills in her analyses. A significant relationship for the Writing Interest Inventory and Writing Traits Scale scores emerged for fourth grade students. Similarly, a significant difference for Writing Traits Scale scores and reading was found for these students. The remaining differences were not significant.

Ney (1977) examined writing miscues among Hispanic and Anglo children. The findings showed no meaningful differences between the groups. Only three Hispanic children were involved in the study. Nelson (1985) described a process based approach to writing for students speaking English as a second language. This college course supported free writing while deemphasizing rules and structure. There were three major segments in the writing experience: (1) drafting, (2) revision and

(3) fine tuning. The instructional approach was grounded in the writer's classroom experience. Grammar and mechanics were studied incidentally.

Galvan (1986) was interested in the influence of the linguistic and cultural background of Spanish-speaking, bilingual-bicultural graduate students on their writing performance. If these factors contribute to writing performance, Galvan may have identified an important variable set which could affect other academic skills as well.

The researcher's study group was made up of ten graduate students who had been educated in Latin America until high school graduation and had resided in the United States for an average of nineteen years. Galvan interviewed and observed his study group and reached the conclusion that their writing was controlled by their acquired language, native language, thought and culture.

Galvan asked the participants to prepare essays on three topics. The first topic was a personal experience, the second, one selected by the investigator and the third, an article which was read earlier. The participants' writing processes indicated that they approached their tasks through three modes, expressive, instrumental and technical. The expressive mode centered on culture, the instrumental, on language, and the technical, on thought. Overall, the participants' writing processes were described as halting, recursive and doubt-ridden.

Procedures

The principal researchers contacted the NAEP by telephone and asked for the prompts used in the Assessment. An NAEP representative said that the prompts were secure and if they could be released for this purpose, the researchers would be notified by mail. No answer came from the NAEP for two weeks and the researchers prepared their own prompts. These prompts were based on those used by the NAEP. Three language arts professionals were asked if, in their opinion, the prompts would produce the same type of response as those used by the NAEP in the Assessment. The professionals said that the study prompts were similar to the NAEP prompts and should produce similar responses.

For informative writing, the students were asked to describe a real scary Halloween night or a perfect day. For persuasive writing, they were asked to write a letter to their principal stating why

students should have afternoon recess and for imaginative writing, they were asked to see themselves turned into Dracula, a puppy dog, a Barbie doll or two local athletes, Juan Samuel or Doctor J.

Principals of four elementary schools with high Hispanic student enrollments were asked if they would like to join the study. The four agreed. Twelve intact fourth grade classrooms were randomly selected from a list of those available. A random numbers table was used for this purpose. The selected teachers were asked if they would assist the principal researchers. All of the teachers agreed.

Research team members offered the prompt to the students who responded. All writing activity took place during the morning and students were given as much time as they needed to complete the assignment. All finished their work within an hour.

Results

Our sample included 240 responses. For imaginative writing, we collected ninety-two samples, for informative writing, seventy-two and for persuasive writing, seventy-six. Table 1 presents summary data. We collected information on gender in our study cohort in order to conduct additional analyses. While the NAEP collected these data, no breakdown by gender within ethnic group was readily available. Therefore, we were unable to make comparisons between the NAEP and the study cohort on this variable.

Table 1

Summary Data: NAEP Sample and Study Sample Characteristics

Writing Exercise	NAEP	Study	Male	Female
Imaginative	162	92	46	46
Informative	92	72	27	35
Persuasive	93	76	48	28

We used chi-square to analyze our data. Our sample distributions and those of the NAEP appear in Table 2. The results of the analyses appear as well. The analyses and results on gender appear in Table 3.

Table 2

Nine Year-Old Hispanic Student Writing Performance - NAEP and Study Cohorts:
Imaginative, Informative and Persuasive Writing

	Score					Mean	N
	0	1	2	3	4		
Format							
Imaginative							
NAEP	21 (8%)	66 (26%)	70 (28%)	5 (2%)	0 (0%)	1.40	162
Sample	0 (0%)	28 (11%)	59 (23%)	5 (2%)	0 (0%)	1.74	92
						Chi-Square = 19.5	
						df = 3, rho = .000	
Informative							
NAEP	4 (2%)	40 (24%)	46 (28%)	2 (1%)	0 (0%)	1.50	92
Sample	10 (6%)	54 (33%)	7 (4%)	1 (1%)	0 (0%)	.98	72
						Chi-Square = 31.7	
						df = 3, rho = .000	
Persuasive							
NAEP	9 (5%)	37 (22%)	30 (18%)	17 (10%)	0 (0%)	1.60	93
Sample	2 (1%)	68 (40%)	6 (4%)	0 (0%)	0 (0%)	1.03	76
						Chi-Square = 45.4	
						df = 3, rho = .000	

Chi-square was significant beyond .001 for the three writing samples. Our sample prepared superior exercises for imaginative writing and the NAEP, for informative and persuasive writing. We included the means for comparison. We analyzed the study cohort performance by gender and found that there were no significant differences for imaginative and informative writing. A significant difference for persuasive writing emerged. The females had higher scores than the males.

Table 3

Nine Year-Old Hispanic Student Writing Performance by Gender- Study Cohort:
Imaginative, Informative andPersuasive Writing

	Score					Mean	N
	0	1	2	3	4		
Format							
Imaginative							
Male	0 (0%)	13 (14%)	31 (34%)	2 (2%)	0 (0%)	1.76	46
Female	0 (0%)	15 (16%)	29 (32%)	2 (2%)	0 (0%)	1.71	46
						Chi-Square = .2	
						df = 2, rho = .90	
Informative							
Male	4 (6%)	20 (32%)	3 (5%)	0 (0%)	0 (0%)	.96	27
Female	5 (8%)	26 (42%)	3 (5%)	1 (2%)	0 (0%)	1.00	35
						Chi-Square = .9	
						df = 3, rho = .83	
Persuasive							
Male	2 (3%)	46 (60%)	0 (0%)	0 (0%)	0 (0%)	.95	48
Female	0 (0%)	23 (30%)	5 (7%)	0 (0%)	0 (0%)	1.18	28
						Chi-Square = 10.1	
						df =2, rho = .006	

Conclusions

Our results showed that significant differences emerged in the three writing formats we examined. In two settings, informative and persuasive, the NAEP sample produced superior writing samples. The study sample performed superior work in the imaginative writing exercise. Within the study sample, we found no significant differences attributable to gender for imaginative or informative writing. There was a significant difference for persuasive writing as the study cohort females produced work which was superior tc the males.

Our results were not consistent. Perhaps Applebee's cautionary note et al.on generalizing from a small sample was verified: The NAEP Hispanic sample was too small and did not represent the

Hispanic population across the country. Additionally, methodological and procedural differences may have contributed to our findings. Our sample was made up of students enrolled in Chapter 1 eligible schools. We did not know the status of the NAEP sample on this variable. The length of time participating Hispanic students spent in the continental United States may differ substantially within the groups and between them. This variable may be influential with regard to writing performance. Finally, we did not have the Assessment's prompts and direct comparisons may not be appropriate because of this difference. Researchers who work in this area ought to consider these variables and take steps to control them through their experimental designs or statistical procedures.

When we examined the students' work, we gained some important insights. Primarily, the students had a great deal of difficulty in working through the imaginative writing task. This difficulty appeared in both cohorts, our sample by observation and the NAEP by comparison. We can only speculate on the cause of this problem; bilingualism, mobility, cultural differences or lack of experience in the writing process may be factors. They ought to be studied.

Reflections

We conducted this study in an attempt to replicate the NAEP's findings among nine year-old Hispanic students. We found that the distribution of scores in our sample was dissimilar from the NAEP's and to a degree, we feel that we achieved our objective. We hope we have shown that the NAEP findings can be used as a criterion in the schools and do not exist for their own sake. We will try to conduct more studies in writing as well as in other NAEP disciplines in an attempt to apply the NAEP's efforts in the classroom. Perhaps researchers can use the system we applied or others in their attempts to work with and extend the NAEP's findings.

References

Casillas, Marian E. "A Study of the Interrelationship among Writing Interest, Selected Writing Traits, and Measures of Reading and Writing of Fourth and Sixth Grade Hispanics." 1986, Dissertation Abstracts International, 1987, 47, 4308A.

Cooper, Charles and Odell, Lee (eds.), Evaluating Writing: Describing, Measuring, Judging, Urbana, IL: National Council of Teachers of English, 1977.

Farmer, Mary. " A Comparison of Process and Traditional Approaches to Large Scale Writing Assessment: Investigating Score Magnitude, Interrater Reliability, and Concurrent Validity." 1986, Dissertation Abstracts International, 1986, 47, 158A.

Fuller, Deborah A. "Evaluation and Curriculum: A Case Study of Primary Trait Scoring." 1985, Dissertation Abstracts International, 1986, 46, 2601A.

Galvan, Max. The Writing Processes of Spanish-Speaking Bilingual/Bicultural Graduate Students.1986. (ERIC Document Reproduction Service No. ED 270 744)

Isern, Margarita, "An Investigation of Bias in Tests of Writing Ability for Bilingual Hispanic College Students." 1986, Dissertation Abstracts International, 1986, 47, 2135A.

Lloyd-Jones, R. "Primary Trait Scoring." in Cooper and Odell (eds.), Evaluating Writing: Describing, Measuring, Judging, Urbana, IL: National Council of Teachers of English, 1977, 33-68.

Mitchell, Karen and Judy Anderson, "Reliability of Holistic Scoring for the MCAT Essay," Educational and Psychological Measurement, Autumn1986, 46, 3, 771-76.

Ney, James W. Miscue Analysis: The Writing of Three Hispanic American Students in a Class of Twenty Fourth Graders. Urbana, IL: National Council of Teachers of English, 1977. (ERIC Document Reproduction Service No. ED 161 077)

Odell, Lee. "Defining and Assessing Competence in Writing." In Charles Cooper (ed.). The Nature and Measurement of Competency in English. Urbana, IL: National Council of Teachers of English, 1981, 95-138.

White, Edward M. Teaching and Assessing Writing. San Francisco: Jossey Bass, 1986.